Machine Learning Algorithm for Stroke Disease Classification

[1]Prof. Vishal Shinde, [2]Mr.Nikhil Patil, [3]Mr.Omkar Sonar, [4]Mr.Nishant Shedage

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharshatra, India.

[1]mailme.vishalshinde@gmail.com, [2]nikhilpatil5965@gmail.com, [3]omkarsonar218@gmail.com, [4]shedagenishant33@gmail.com

Abstract- Stroke is the most significant element of death as well as obesity in various countries. This study analyzes CT scan image enhancement data for stroke patients by improving image quality to improve image effects and noise reduction, and using ML algorithms to separate patient images into two subtypes of stroke i.e. ischemic stroke and stroke haemorrhage. Eight ML algorithms are operated in this study to classify stroke i.e., Naive Bayes, K-Nearest Neighbors, Logistic Regression, Random Forest, Decision Tree, Multi-layer Perceptron (MLP-NN), Vector Support and Deep Learning Machine. Our results indicate that Random Forest produces the maximum level of precision (95.97%), as well as accurate values (94.39%), recall values (96.12%) and f1-Measures (95.39%).

Keywords:- CT scan , Machine Learning Algorithms, Stroke.

I. INTRODUCTION

Stroke is a prime reason of mortality as well as high morbidity in many lands. It is necessary to get a proper diagnosis before the start of stroke treatment, because the treatment of stroke build upon on the category of stroke affected. This study classifies stroke patients into ischemic stroke and hemorrhage stroke based on CT scan image data. Ischemic stroke is usually occur due to blockage of artery. While hemorrhage is caused by bleeding of brain tissue. Research connected to diagnosis and prognosis of stroke is conducted by Chiun-Li Chin, et al. [4] develops an automated ischemic stroke diagnostic system using the Deep Learning algorithm. The CNN architecture utilized two conversion layers, namely the integration layer and the fully integrated layer. The prime reason of the floor integration layer, which means that the layer will compress the amount of data and parameters to reduce the problem of overfilling. Separation results obtained 90% accuracy. Other researchers using CNN's Deep Learning architecture are Marbun, JT. et al. [5] used the database for open stroke found on it www.radiopaedia.org to classify patient data into three classes: common stroke, ischemic, and hemorrhagic stroke using CT scans. The accuracy obtained is 90%. While other researchers used the same database, Badriyah, Tessy et al. [6] performs hyperparameter optimizations in Deep Learning algorithm to improve the accuracy of stroke diagnosis. Random search algorithm and Bayesian search hyperparameter tuning in Advanced Reading can increase accuracy up to 100%. The SVM algorithm also provided satisfactory results in a study conducted by Jenna R.S and Drs. Sukesh Kumar [9] predicted strokes using various kernel operations on the Vector Support Machine path. Leading test results obtained from Kernel Linear Function at 91.7% and Polynomial at 89.0%. The Study Concluded the significance of the phenotypic form of classifications of stroke; it also describes of its reliability in accuracy.

II. AIMS AND OBJECTIVE

a) Aim

Aim is to investigate machine learning based technique for patient disease forecasting by prediction result in comparing best accuracy with evaluation of GUI application.

b) Objective

- Stroke disease is Major reason of mortality among all other disease.
- Because of a deficiency of resources in medical field. The Classification of stroke sometimes may be problem.
- Target Prevention on patient with high risk for developing a stroke.
- The Machine learning technology can be very well adpted to do prediction of stroke disease.
- Help physicians to take proactive health.

III. LITERATURE SURVEY

Paper 1: Burden of Ischemic and Hemorrhagic Stroke

Global stroke epidemiology is developing speedily. Even if age-standardized rates of stroke deaths have declined

 DOI : 10.35291/2454-9150.2022.0170

worldwide in last 20 years, the total numbers of peoples who have a stroke all year, as well as live with the effects of stroke or else die from their stroke, are increasing. Daily updates on the existing type of stroke burden significant for improving our understanding on stroke epidemiology as well as facilitate organization as well as planning of confirmation-based stroke care. Objectives: This study goal to calculate incidence, prevalence, deaths, disability-adapted life years (DALYs) as well as years lived along disability (YLDs) along with their trends to ischemic stroke (IS) as well as hemorrhagic stroke (HS) to 188 countries since 1990 to 2013. Methodology: Stroke incidence along with prevalence, deaths, DALYs as well as YLDs were calculated using each available data on deaths as well as stroke incidence, prevalence including excess mortality.

Conclusion: The global strokes burden continue to increase worldwide Strong measures to prevent and control strokes are immediately needed to prevent and eventually invert the stroke epidemic, although universal entrance to stroke service should be a preference.

Paper 2: Stroke Epidemiology in South, East, and South-East Asia: A Review

Asia, that holds 60% of people worldwide, comprises some growing countries that are in economic development. This paper review epidemiology of stroke within South, East as well as South-East Asia. Data on the epidemiology of stroke inside South and East, as well as South-East Asia were received through Global Burden of Diseases study (deaths, disability-adapted life-years [DALYs] lost by cause of stroke), WHO (vascular risk elements in the society), as well as publications in PubMed. Age- as well as sex-standardized death is the lower in Japan, as well as higher in Mongolia. Society-based incidence data of just a some countries are available, alongside the lower rates discovered in Malaysia, with the higher in Japan including Taiwan. The accessibility of prevalence data is on top of incidence data, however different study technique were applied for case-finding, alongside various age bands.There are differences in the epidemiology among countries inside South, East, also South-East Asia. additional study on stroke burden is needed.

Paper 3: Automated Ischemic Stroke Subtyping Based on Machine Learning Approach

Ischemic stroke sub-typing was not just precious for effcient intervention with treatment, equally significant to the prediction of ischemic stroke. Manual judgement of disease classifications was timeconsuming, error-prone, also limits scaling for big datasets. In this reasearch, an integrated ML approach was applied to classify the type of ischemic stroke on The International Stroke Trial (IST) dataset. We studied the ordinary problems of feature selection including prediction in medical datasets. primarily, the significance of features were ranked by the Shapiro-Wilk algorithm including Pearson correlations among features were examined. Next, we utilized Recursive Feature Elimination trough Cross-Validation (RFECV), that integrated linear SVC with Random-Forest-Classifier along with Extra-Tree-Classifier and AdaBoost-Classifier, including Multinomial-Naïve-Bayes-Classifier such as estimator, to choose robust features significant to ischemic stroke subtyping.

IV. EXISTING SYSTEM

Stroke is a term used by most health professionals to describe brain aa well as spinal cord injuries caused by poor blood supply. The Stroke project has its own definition based on different perspectives; Globally, however, the stroke triggers a clear visceral response. The current effective research program has its own impact only on the class of ischemic stroke (IS); the change of hazard factors on stroke and its classification is not allow due importance in research; and most of the research work classified the stroke with the help of only two or three ML algorithms, and stroke classification using aggregated data received from case pages and case summaries was not attempted. In this study, mining strategies are present to overcome the barriers listed above and to distinguish the type of strokes accurately.

DISADVANTAGE OF EXISTING SYSTEM:

- Download the processed data using the stemming novel algorithm and avoid conflicts related to word size and capture errors found in the system.
- For classify a type of stroke with reasonable accuracy by providing importance to the variation in the dataset.
- Most of the existing method works with patient case sheet data not for images data.

V. PROBLEM STATEMENT

Diagnosis of stroke is an important and complex medical procedure. Medical examination is considered an important but complex task that wants to be done accurately and effectively the automation of this system can be very beneficial. Existing system method is less accurate. A limited database used in this system.

VI. SCOPE

In this Project, stroke disease classification is implemented by eight various algorithms of Machine Learning namely K-Nearest Neighbors, Logistic Regression, Naive Bayes, Random Forest, Decision Tree, Deep Learning, Support Vector Machine and Multi-layer Perceptron (MLP-NN).

· This project will classify which stroke the patient has had.

· It is improve the images quality of CT scans of stroke patients by optimizing the quality of image to improve image results.

· More accurate diagnosis of this disease by doctors.

 DOI : 10.35291/2454-9150.2022.0170

VII. COMPARTIVE STUDY

Sr. No.	Author	Project Title	Publication	Method	Purpose
1.	Feigin V.L, Parmar P, Krishnmurthi R.V, Norrving B, Mensah G.A	Burden of Ischemic and Hemorrhagic Stroke.	Neuroepidemiology, 2015	Stroke prevalence, occurrence, humanity, DALYs and YALDs were estimated operating all accessible data on humanity and stroke incidence and excess mortality.	This study goals to estimate incidence, occurrence, humanity, disability-adjusted life years (DALYs) and years survived with incapacity (YLDs) and their leanings to ischemic stroke (IS) as well as hemorrhagic stroke for 188 countries from 1990 to 2013.
2.	N. Venketasubramanian, B.W. Yoon, J. Pandian, J.C.Navarro	Stroke Epidemiology in South, East, and South-East Asia: A Review	J Stroke,2017	Data on stroke mortality and morbidity were obtained from the GBD study, and vascular risk factors in the community were obtained from the WHO database. All the data search was conducted by a single author (NV). The data were then tabulated, stratified according to geographical regions.	This study continued conducted to analysis the fresh epidemiology of the stroke in South, and South-East Asia, as well as mortality, occurrence, occurrence, DALYs lost, stroke subtype or danger issues, based on statistics since the Global Burden of Disease (GBD) study, (WHO) World Health Organization and the recent main publications from Asian republics 3.
3.	Peng Xu, Gang Fang, Wenbin Liu	Automated Ischemic Stroke Subtyping Based on Machine Learning Approach	IEEE,2020	Used Recursive Feature Elimination with Cross-Validation, Which incorporated linear SVC, Random-Forest-Classifier, Extra-Tress-Classifier, AdaBoost-Classifier, and Multinomial-Naïve-Bayes-Classifier.	This study, an combined machine learning method was used to categorize the subtype a ischemic stroke. The (IST) International Stroke Trial dataset.This Paper well thought-out the usual difficulties of feature selection and calculation in medical datasets.

VIII. PROPOSED SYSTEM

Stroke is a main reason of mortality as well as high morbidity in causing disability in many countries. It is significant to receive accurate diagnosis prior to stroke treatment begins, as long as treatment for stroke build upon the sort of stroke suffered. This study classifies stroke patients into ischemic stroke and hemorrhage stroke based on CT scan image data. Ischemic stroke is normally occur due to blockage in a blood vessel. While hemorrhage stroke is occured due to bleeding in brain tissue. CNN architecture use uses two convolutional layers, namely the pooling layer as well as the fully connected layer. The motive of the pooling layer is down-sampling, which means the layer will condense the amount of data as well as parameters to decrease the issue of overfitting. Eight Algorithms used in this study to predict accurate results.

ADVANTAGES OF PROPOSED SYSTEM:

- The SVM algorithm also gave satisfying results.
- Data Collection utilized in this study is the CT scan data of patients suffers with stroke obtained from web scrapping.
- Gray Level Co-occurrence Matrix (GLCM) method. The features that GLCM has are 6 features like contrast,dissimilarity,homogeneity, correlation, ASM, energy.

Algorithm: Naïve Bayes, K-Nearest Neighbors, Logistic Regression, Random Forest, Decision Tree, Deep Learning, Multi-layer Perceptron (MLP-NN).

IX. ALGORITHM

The general idea of working of proposed system algorithm is given as follow:

Step 1: Start
Step 2: Loading Dataset of Image
Filepath ← settings.MEDIA_ROOT + "\\" + filename
Step 3: Extract The Image Data
PATCH_SIZE ← 20
image ← cv2.imread(filepath, 0)
Step 4: Test and run Mode
Dataset ← GLCMtableModel.object.all()
df ← read_frame(dataset)
X ← df.iloc[:,1:7].values
y ← df.iloc[:,-1].values
X_train,X_test,y_train,y_test←train_test_split(X,y, test_size ← 0.25,random_state ← 0)
Step 5: Logisticregression implementation
reg ← LogisticRegression()

 DOI : 10.35291/2454-9150.2022.0170

```
reg.fit(self.X_train,self.y_train)
y_pred ← reg.predict(self.X_test)
lg_acc ← accuracy_score(self.y_test,y_pred)
lg_precc←precision_score(self.y_test,y_pred)
lg_recall ← recall_score(self.y_test,y_pred)
lg_f1Scr ← f1_score(self.y_test,y_pred)
cm ← confusion_matrix(self.y_test,y_pred)
```

Step 6: KNN implementation

```
knn ← KNeighborsClassifier()
knn.fit(self.X_train, self.y_train)
y_pred ← knn.predict(self.X_test)
knn_acc←accuracy_score(self.y_test,y_pred)
knn_precc←precision_score(self.y_test,y_pred)
knn_recall←recall_score(self.y_test,y_pred) knn_f1Scr ←
f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)
```

Step 7: Naivebayes implementation

```
nb ← GaussianNB()
nb.fit(self.X_train, self.y_train)
y_pred ← nb.predict(self.X_test)
nb_acc←accuracy_score(self.y_test, y_pred)
nb_precc←precision_score(self.y_test,y_pred)
nb_recall← recall_score(self.y_test, y_pred)
nb_f1Scr ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)
```

Step 8: DecisionTree implementation

```
dt ← DecisionTreeClassifier()
dt.fit(self.X_train, self.y_train)
y_pred ← dt.predict(self.X_test)
dt_acc←accuracy_score(self.y_test, y_pred)
dt_precc←precision_score(self.y_test,y_pred)
dt_recall ← recall_score(self.y_test, y_pred)
dt_f1Scr qq ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)
```

Step 9: RandomForest implementation

```
rf ← RandomForestClassifier()
rf.fit(self.X_train, self.y_train)
y_pred ← rf.predict(self.X_test)
rf_acc← accuracy_score(self.y_test, y_pred)
rf_precc←precision_score(self.y_test,y_pred)
rf_recall← recall_score(self.y_test, y_pred)
rf_f1Scr ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)
```

Step 10: SVM implementation

```
cls ← SVC()
cls.fit(self.X_train, self.y_train)
y_pred ← cls.predict(self.X_test)
svm_acc←accuracy_score(self.y_test,y_pred)
svm_precc←precision_score(self.y_test,y_pred)
svm_recall←recall_score(self.y_test,y_pred)
svm_f1Scr ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)
```

Step11: MultiLayerPerceptronimplementation

```
cls ← MLPClassifier(activation ← 'relu')
cls.fit(self.X_train, self.y_train)
y_pred ← cls.predict(self.X_test)
mlp_acc←accuracy_score(self.y_test,y_pred)
mlp_precc←precision_score(self.y_test,y_pred)
mlp_recall←recall_score(self.y_test,y_pred)
mlp_f1Scr ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)
```

Step 12: DeepLearning implementation

```
classifier ← Sequential()
classifier.add(Dense(output_dim ← 6, init ← 'uniform', activation ←
'relu', input_dim ←6))
classifier.add(Dense(output_dim ← 6, init ← 'uniform', activation
← 'relu'))
classifier.add(Dense(output_dim ← 1, init ← 'uniform', activation
← 'sigmoid'))
classifier.compile(optimizer ← 'adam', loss ← 'binary_crossentropy',
metrics ← ['accuracy'])
classifier.fit(self.X_train,self.y_train,
batch_size ← 10, nb_epoch ← 100)
y_pred ← classifier.predict(self.X_test)
y_pred ← (y_pred > 0.5)
cm ← confusion_matrix(self.y_test, y_pred)
dl_acc← accuracy_score(self.y_test, y_pred)
dl_precc←precision_score(self.y_test,y_pred)
dl_recall←recall_score(self.y_test, y_pred)
dl_f1Scr ← f1_score(self.y_test, y_pred)
```

Step.10: Exit

X. MATHEMATICAL MODEL

1. Classification

Classification Technique precisely predicts target class with regards to all given data Pre-processed data is integrated into various classification algorithms in order to measure accuracy of all classification method. The ROC to represent the accuracy in regard to categories, operating character (ROC) curves used in this function. In the case of dichotomic segments they are predicted using continuous variable x mean value " calculated for each event. Based on threshold T, conditions are divided when X> T, then positive; otherwise, it is negative.

$f_1(x)$ the potential for congestion work that X can follow if the model says yes. f_0 the potential for congestion work that X can follow if the model is negative. Therefore, true confirmation values and false values are provided:

True positive rate (T) = $\int_T^\infty f_1(x)dx$ .....(2)

False positive rate (T) = $\int_T^\infty f_0(x)dx$ .....(3)

Hence, ROC curves plot the true positive rate versus the false positive rate with threshold T as the varying parameter.

2. ANN

ANN is constructed along 22 inputs as well as one unseen layer along ten neurons as wll as two output (IS and HE). The network is trained with the stochastics gradient descent algorithm. Stochastic gradient descent or additional gradient descent are stochastic approximations of gradient descent optimization. It also helps in optimizing differentiable objective function through an iterative method. It finds maxima or minima by iteration. The next is the pseudocode for stochastic gradient descent:

1. starting vector parameters w and learning rate g is chosen

2. Repeat the coming steps until an approximate minimum is obtained:

- Examples are randomly shuffled in the training set
- For i = 1 to n, do

$w = w - \eta\, sQ_i(w)$

where Q(w) represents the empirical risk and Q_i (- w) represents the value of the loss function at ith example From

 DOI : 10.35291/2454-9150.2022.0170

the samples, 300 samples were used for training and the rest 207 sample were utilized for testing. Implementing the neural network achieved 95% classifications accuracy along with standard deviation of 14.69.

3. SVM

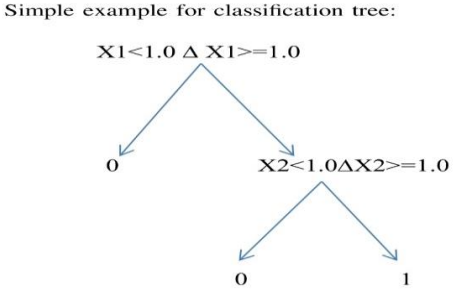
SVM is other classification method utilized for predicting stroke, which was prepared with statistical knowledging theory is widely used in many domains, from image identification to bioinformatics. SVM training algorithm build a prototype that allocate new entities to current group either creates a latest group. SVM that is labeled as supervised ML is utilized in favor of both regression challenges as well as classification. Algorithm plots every entity from the datasets in n-dimensional space (n-number of feature) as a point. All feature's rate is considered as rate of specific coordinate. Classification is carried out by findingout hyperplane that differentiate both the classes. Differentiating function in SVM is explained as a linear mixture of kernels connected through support vectors

$$f(x) = \sum_{x\,\epsilon S} \propto_j y_j K(x_j, x) + b \quad (4)$$

In Eq. (4), x_j indicates the patterns of the training set, $y_j \epsilon \{+1, -1\}$ indicates the respective class labels, and S indicates a set of support vectors [32, 33]. The pre-process samples are classified using several kernels. This methods, linear SVM, medium Gaussian SVM, and coarse Gaussian SVM produced the highest accuracy of 91.5%. ROC curves received for the various SVM kernels accomplished the higher accuracy along a training time of 2.28 s, which are depicted.

4. Decision Tree

Decision tree constructs a tree structure using classification either regression models. Decision tree is created by splitting the datasets into small subsets. Tree classifie the dataset, however it doesn't know to determine on itself with example of patient. Every single dataset falls under any of the labeled class. As a result, it comes under the point of view of supervised learning instead of unsupervised learning. Tree is construct using the obtain information, with prolonged development are projected with a sole pruning methodology. It classifies every types of data, for example continuous, discrete, concise, along with easy to infer, as well as the throughput is ever a humans-readable.



The prediction begin at the head node of the tree (r) as well as examine the decision alongside the primary predictor (X1): If its value is lower than 1.0, later it goes to left branch as well as the tree classifie it prediction such as type 0; otherwise it goes to right branch which is again a prediction is done on the basis of the second predictor X2. X2 value is lower than 1.0, so it goes the left branch including the tree classifie it prediction such as type 0; otherwise it goes the right branch as well as the tree classifis it prediction such as type 1.The tree generate accuracy through kernels as simple trees, medium trees, including complex trees, that are shown in Fig. 5, 6, and 7. The basic tree generate the highe accuracy (90.7%) in comparision to the other through a training time of 1.45 s.

5. Logistic Regression

Logistic regression (LR) is based on predictive analysis, which describes the data and provides the relationship between independent variables and dependent (binary) variables. For example, does the patient's age, hypertension, and diabetes mellitus level impact the stroke patient (yes or no)? The outcome of the process is either 0 or 1, which is tagged as dependent, and other predictors are taken as covariates. This methodology is used in various fields, including machine learning to predict the presence of disease (based on factors) and in the marketing field. In this work, the LR method produced an accuracy of 90.6% with a training time of 8.51 s; when the sample data were fed in, 10-fold cross-validation is done.

$$i = \text{Logistic regression}\,(p) = 1n\left(\frac{p}{1-p}\right).$$

6. Naïve Bayes

Naïve Bayes classifier is a classifier based on Bayes' theorem. It is a special class of Machine Learning classification algorithms. Bayes claimed that the global is not uncertain or not probabilistic, but instead of that we study from the global through approximations, that make us gets closer and closer near the truth the more evidence we have. Naïve Bayes classifier assumes that the presence or absence of an attribute is not probabilistically related to the presence or absence of another attribute, contrary to what happens in the real world. Naïve Bayes classifier allows easily built probability-based models with excellent performance, due to its simplicity. The Naïve Bayes classifier or algorithm converts the data set into a frequency table. A probability table is created in order for the various events to occur. Naïve Bayes is applied to calculate the posterior probability of each class and the class of the prediction is the class with the highest probability.

$$P(A|C) = \frac{P(C|B)\,P(B)}{P(C)}.$$

 DOI : 10.35291/2454-9150.2022.0170

7. KNN

The K Nearest Neighbour (K-NN) classifier is a supervised learning algorithm . The idea of the classifier is very intuitive. K-NN algorithm classifie each new data into the class of its nearest neighbor. It calculate the distance from the latest element to all of the existing one then order these distances from small to larget to select the class to belong to. This class will therefore be the one with the high frequency with shortest distances. The K-NN algorithm is broadly used for pattern classification .

$$d(x, x') = \sqrt{(x_1 - x_1')^2 + \ldots + (x_n - x_n')^2}$$

8. Radom Forest

Ensemble method is procedure that combine the prediction from various ML algorithms to make predictions that are more accurate than some individual model. Bagging algorithms create an ensemble of model (classifiers or else predictors), that is learning schemes where all models gives an similar-weighted predictions. The method use in bagging classifiers type is a random forest.random forest algorithm operate in two stage. In the primary stage, creations of random forest is conducted, succeeded by a predictions with random forest classifiers that was made in the first stage . The following is the pseudo-code for the creation of random forest:

1. Random selection of "m" features from the total "n" features, where m << n

2. Among the "m" features, the node "x" calculation is done by means of best split point

3. Node is divided into children node using the best split

4. Repeat the steps 1-3 until "y" number of nodes are reached

5. Repeat the steps 1-4 for "N" number of times to build forest as well as to create "n" number of nodes

The following is the pseudo-code for prediction through random forest classifier (first stage):

1. Store the predicted outcome as the target by using the rules as well as test features of a randomly made decision tree

2. Predicted target's votes are calculated

3. The expected target along high vote is studied as the final predictions within random forest algorithms Boosting algorithm create an ensemble of classifies; all gives a weighted vote. The method utilized in the boosting classifiers type is AdaBoost. AdaBoost indicate a method of training a boosted classifiera, as well as it takes the below form

$$F_T(x) = \sum_{t=1}^{T} f_t(x) \qquad (5)$$

where f_t is indicated as a poor learner which takes x as the input object and it return the class of the object as a value. Tth classifier indicate positive in case the objects is in positive class or it comes under negative class.

The next is the pseudocode of AdaBoost:

1. Primary weight value, w_i = 1/n (n, represents the no of total observations) is assigned to each observation, X_i

2. "Weak" model is trained (often a decision tree)

3. In each observation,

3:1. W_i is increased, for incorrect predictions

3:2. W_i is decreased, for correct predictions

4. A poor model is train by giving extra priority to high weighs in the observation.

XI. SYSTEM ARCHITECTURE

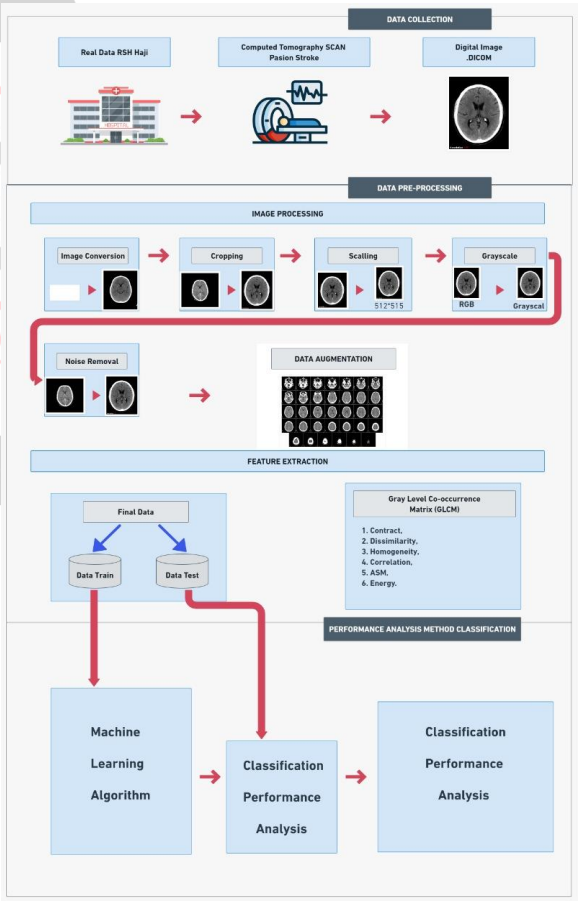


Fig.1: System Architecture

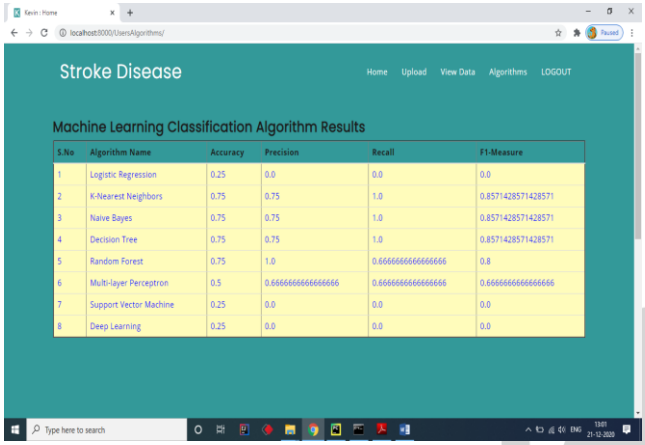
XII. ADVANATGES

- The SVM algorithm also gave satisfying results.

 DOI : 10.35291/2454-9150.2022.0170

- Data Collection utilized in this study is the CT scan data of patients suffers with stroke obtained from web scrapping.
- Gray Level Co-occurrence Matrix (GLCM) method. the features that GLCM has are 6 features including contrast, dissimilarity, homogeneity, correlation, ASM, energy.
- To reduce doctor risk.
- High Performance
- The Cost of medication will be minimized.

XIII. DESIGN DETAILS



XIV. CONCLUSION

Thus, We have tried to implement the paper "Tessy Badriyah, Nur Sakinah, Iwan Syarif, and Daisy Rahmania Syarif, Fellow", "*Machine Learning Algorithm for Stroke Disease Classification*", IEEE 2020 and according to the implementation the conclusion is for Stroke Disease Classification This study uses machine learning algorithms to classify stroke data on CT Scan image data. Before classification, image processing and feature extraction are performed on image data. And after that the comparison of 8 (eight) algorithms is used to do the classification, namely: Naive Bayes, K-Nearest Neighbors, Logistic Regression, Random Forest, Decision Tree, Deep Learning Multi-layer Perceptron (MLP-NN)and Support Vector Machine. Based on our experiment, classification algorithm using the Random Forest method has the high standard of accuracy compared to other tested classification algorithms. However the accuracy of classification method using optimization default parameter value has not been carried out. Henceforth there is a possibility for the classification model to be improved to achieve. Parameter tuning needs to be done to machine learning algorithm used in order to increase its accuracy.

REFERENCE

[1] " Tessy Badriyah ", " Nur Sakinah ", "Iwan Syarif ", and " Daisy Rahmania Syarif" , et al. "Machine Learning Algorithm for Stroke Disease Classification", IEEE 2020.

[2] V. L. Feigin et al., Update on the global burden of ischemic and hemorrhagic stroke in 1990-2013: The GBD 2013 study, vol. 45, no. 3. 2015.

[3] N. Venketasubramanian, B. W. Yoon, J. Pandian, and J. C. Navarro, Stroke Epidemiology in South, East, and South-East Asia: A Review, vol. 20, no. 1. 2018.

[4] G. A. P. Singh and P. K. Gupta, Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans, vol. 3456789. Springer London, 2018.

[5] C. L. Chin et al., An automated early ischemic stroke detection system using CNN deep learning algorithm, vol. 2018-Janua, no. iCAST. 2018.

[6] J. T. Marbun, Seniman, and U. Andayani, Classification of stroke disease using convolutional neural network, vol. 978, no. 1. 2018.

 DOI : 10.35291/2454-9150.2022.0170